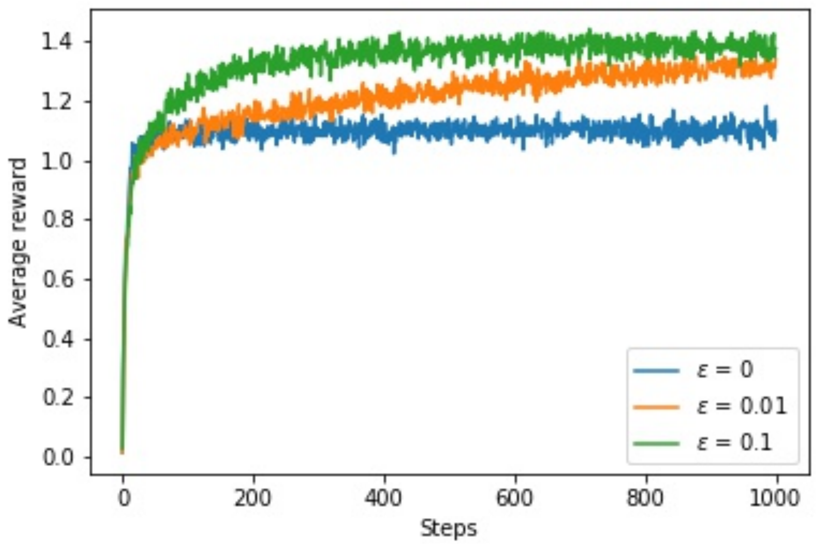
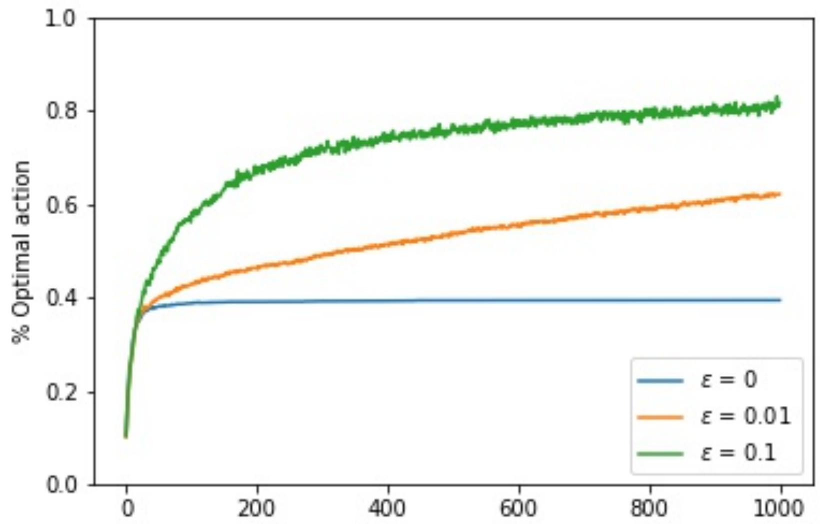
RL Assignment 1

Srivatsava Kesanupalli MT18054

Exercise 2.5 - The 10-armed bandit problem for $\boldsymbol{\varepsilon}$ = 0, 0.01, 0.1



Average reward



Optimal action

Figure 2.3 - $Q_1 = 5$ with $\boldsymbol{\varepsilon} = 0$ vs. $Q_1 = 0$ with $\boldsymbol{\varepsilon} = 0.1$

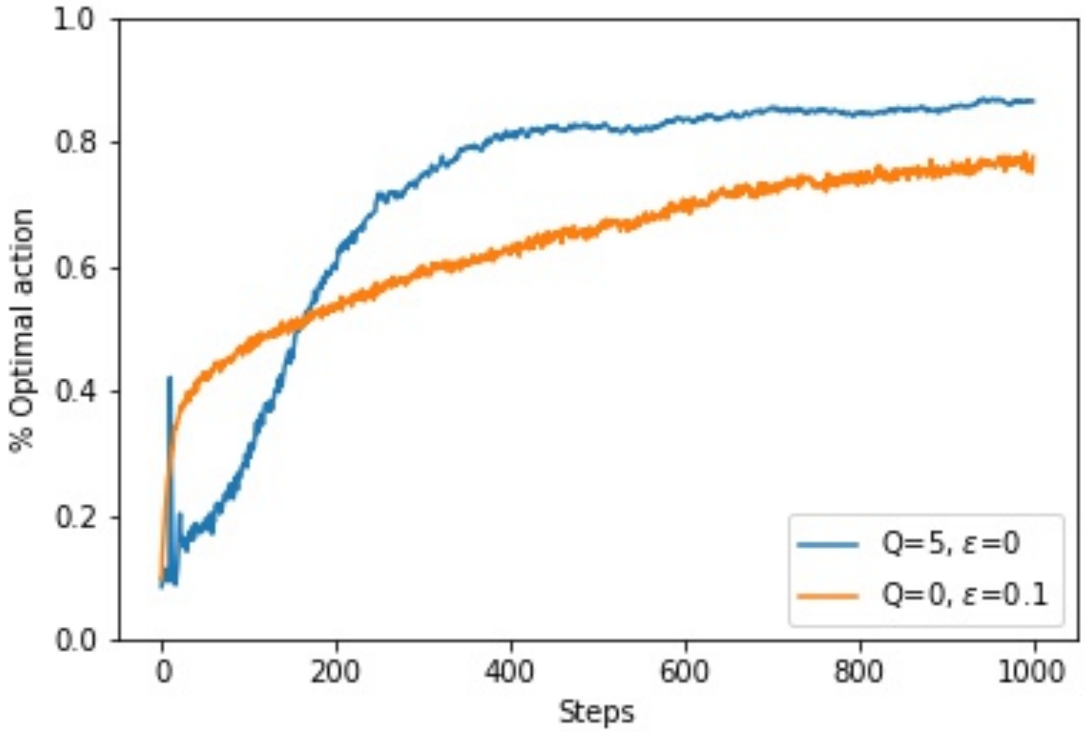
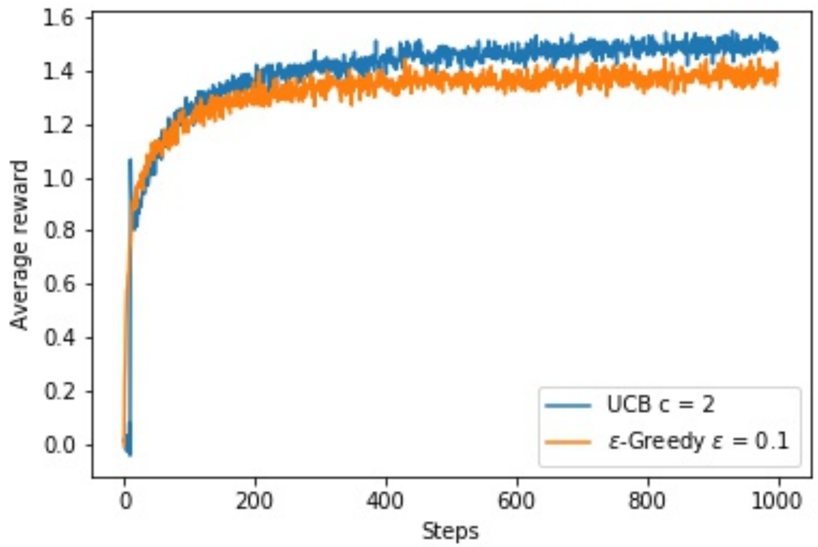
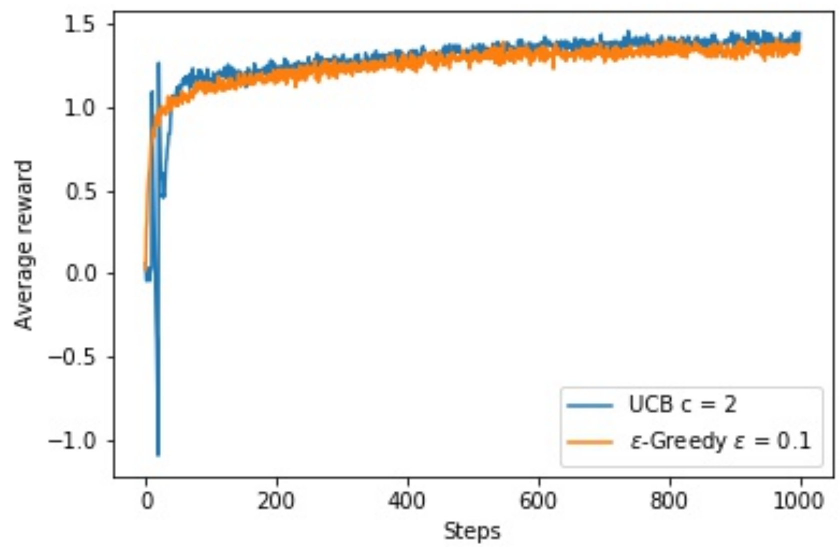


Fig 2.3

UCB vs. ε-Greedy



UCB stationary vs ε-Greedy



UCB non-stationary vs ε-Greedy

UCB performs better than ε-Greedy as the error term in the estimate penalises for not exploring an action. The performance can be seen for both stationary as well as non-stationary cases.